

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:36 ; Search time 170.72 Seconds
(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631a-5_COPY_33_75

Perfect score: 248
Sequence: 1 NQEDPQTECCQCORRCRQEE.....RQOYCCORCKEICEEEY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	44.4	566	2 S22477	vicillin precursor
2	105	42.3	588	1 FMCNAB	alpha-globulin B p
3	104	41.9	509	2 S08059	alpha-globulin typ
4	96	38.7	411	2 T29475	hypothetical prote
5	94	37.9	605	2 S06398	alpha-globulin typ
6	78.5	31.7	425	2 T18592	hypothetical prote
7	78.5	31.7	600	2 T18593	hypothetical prote
8	77	31.0	572	2 T29880	hypothetical prote
9	77	31.0	1513	2 T23681	hypothetical prote
10	75.5	30.4	242	2 T29699	hypothetical prote
11	75	30.2	388	2 T31887	hypothetical prote
12	75	30.2	388	2 T31888	hypothetical prote
13	74	29.8	438	2 T31889	hypothetical prote
14	74	29.8	445	2 T31898	hypothetical prote
15	73.5	29.6	810	2 T44430	protein PVI00 [imp
16	73	29.4	330	2 T25169	hypothetical prote
17	72.5	29.2	335	2 T31559	hypothetical prote
18	71.5	28.8	335	2 T31560	hypothetical prote
19	71.5	28.8	335	2 T31561	hypothetical prote
20	71	28.6	47	2 JCS557	arginine/glutamate
21	68.5	27.6	356	2 S31574	hypothetical prote
22	68	27.4	154	2 T27967	hypothetical prote
23	67.5	27.2	314	2 T27886	hypothetical prote
24	66.5	26.8	273	2 T20930	hypothetical prote
25	66.5	26.8	637	2 S35221	globulin Bgl prec
26	66.5	26.8	654	2 T30136	hypothetical prote
27	66	26.6	111	2 T20452	hypothetical prote
28	65.5	26.4	33	2 A41822	antimicrobial pept
29	65	26.2	419	2 T18450	hypothetical prote

30	65	26.2	910	2 A34721	androgen receptor
31	65	26.2	911	2 B34721	androgen receptor
32	65	26.2	919	2 A39248	androgen receptor
33	65	26.2	1306	2 T13592	hypothetical prote
34	64	25.8	648	1 T01150	protein kinase (EC
35	63.5	25.6	51	1 HSMSS1	protamine - mouse
36	63.5	25.6	51	2 S03997	protamine 1 - rat
37	63	25.4	1390	2 T14004	histidine kinase h
38	62	25.0	1969	2 T08875	trifla protein - sl
39	62	25.0	4957	2 T03455	ALR protein - huma
40	62	25.0	5262	2 T03454	ALR protein - huma
41	61	24.6	223	1 KASHS2	alpha-s2-casein pr
42	61	24.6	429	2 S29565	apolipoprotein A-I
43	60	24.2	285	1 I46207	involucrin - dog
44	60	24.2	1038	2 T02634	rep protein homolo
45	60	24.2	1094	2 S49313	protein kinase - s

ALIGNMENTS

```

RESULT 1
S22477
vicillin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L.; Fritze, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of coco
A:Reference number: S22477; MIDID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCH>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-566/Product: vicillin #status predicted <MAT>

Query Match 44.4%; Score 110; DB 2; Length 566;
Best Local Similarity 47.5%; Pred. No. 0.00028;
Matches 19; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 2 QEDPTECCQCORRCRQEESDPQQOYCCORCKEICEEE 41
DB 78 EEDLRQYQCGRCQEDQGGREDDQCCORCKWEYKQEE 117

RESULT 2
FMCNAB
alpha-globulin B precursor (clone C72) - upland cotton
W:Alternate names: seed storage protein; vicillin precursor
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911
R:Chian, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA
A:Residues: 1-588 <CHI>
A:Cross-references: GB:M16891; NID:q167374; PIDN:AAA33071.1; PID:q167375
A:Experimental source: var. Coker 201
R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

```

Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed
A:Reference number: S06398
A:Accession: S06911
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-25/domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage preprotein #status predicted <MAT>
E:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	42.38;	Score 105;	DB 1;	Length 588;
Best Local Similarity	47.58;	Pred. No. 0.00093;		
Matches 19;	Conservative 9;	Mismatches 10;	Indels 2;	Gaps 1.

```

Qy      3 EDPOTECCQCCRCRQDES DP RQOQYCQRCKEICEEEE 42
        ||| :||: ||| | :|: || :|: :|:
Db      81 EDPQRVEECQECRQEE - RGRPQCQCCCLKRFEEQO 118

```

```

RESULT      3
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borrolo, K.; Kamalay, J.A.; Dure III, L.
Plant MOL. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed
A:Reference number: S06398
A:Accession: S08059
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-509 <CHL>
C:Superfamily: glycinin

```

Query Match	41.98;	Score 104;	DB 2;	Length 509;
Best Local Similarity	48.78;	Pred. No. 0.0011;		
Matches 19;	Conservative 8;	Mismatches 10;	Indels 2;	Gaps 1

```
QY      4 DPQTECCQCCORRCROESDPDROQYCCORCKEICEEEEE 42
      III :III: IIIII III II:II : I:I:
Db      1 DPQRYEECCQECRQEE--RQDPQCCQRCCLKREQEQQ 37
```

RESULT 4
T29A75
hypothetical protein T01D1.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29A75
R:Bradshaw, H.; Wohlmann, P.
submitted to the EMBL data library, November 1996
A:Description: The sequence of C. elegans cosmid T01D1.
A:Reference number: Z20623

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-411

A:Cross-references: EMBL:U80455; PIDN:AAB37887.1; GSPDB:GN00020; CESP:T01D1.6
A:Experimental source: strain Bristol N2; clone T01D1
C:Genetics:
A:Gene: CESP:T01D1.6
A:Map position: 2
A:Introns: 25/3, 304/3
A:Superfamily: glialdin

Query Match	38.7%	Score 96;	DB 2;	Length 411;
Best Local Similarity	39.5%;	Pred. No. 0.0059;		
Matches 17;	Conservative 11;	Mismatches 11;	Indels 4;	Caps 2

```

Qy      2 QEDPQTECCQ-CQRCRCQDESPDQOQYICRRCHEICEEEFY 43
      | : | : | | | : | : | : | : | : | : | : | : |
Db     138 QQPAPQCCQCDQCAACPQQQ---QPQQCCQCCQCTTCQSDQY 17

```

RESULT	5
S06398	

alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence, revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398
R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX
A:Reference number: S06398
A:Accession: S06398
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-605 <CHL>
C:Superfamily: glycinin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match 37.9%; Score 94; DB 2; Length 605;
Best Local Similarity 35.9%; Pred. No. 0.012;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Caps 0

```

Oy      2 QEDPQTECQCQCRCRCQESDPRQOQYQCGRCKEICEEE 40
      | : | : ::|||: | | : : | : | : | : | : | :
Db      115 QQPDPKQKFKECQGRQWQEQGRPRKQKQCVKCEKCRQYQED 15

```

```

RESULT      6
T18592
hypothetical protein AC3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18592
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z18995
A:Accession: T18592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WLL>
A:Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
A:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.3
A:Map position: 5
A:introns: 18/3
C:Superfamily: gliadin

```

Query Match	31.7%;	Score 78.5;	DB 2;	Length 425;
Best Local Similarity	27.9%;	Pred. No. 0.36;		
Matches 17;	Conservative	9;	Mismatches 12;	Indels 23;
				Gaps 2

```
Qy      2 QEDPTECQCQRRCR-----QEESDP-----RQQQYORCKEICE 38
        | : ||||| | : ||||| | : ||||| | : ||||| | : ||||| | :
Db    255 QQPSAPQCQCQONTCCQAAPVCQQQGCAPQCQCQGCSPACQCQCSTSCQTQCQCQCCTPCQC 314
```

QY	39	E	39
Db	315	Q	315

A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match	30.28;	Score 75;	DB 2;	Length 388;
Best Local Similarity	37.58;	Pred. No. 0.77;		
Matches	15;	Conservative	7;	Mismatches 16;
				Indels 2;
				Gaps 2

Qy 2 QEDPTEC-QQCRCRQGESDPRQ-QYQRCKEICEE 39
 | : : ||| | || : : : :||:
Db 194 QAQCPCQCQCQCSSCVCVQQQNSNCEPAANTQCSDICQ 233

RESULT 12
T3188
hypothetical protein C03A7.7 - *Caenorhabditis elegans*

A:Cross-references: EMBL:AF016451; PIR:AAH65996.1; GSPD:GN00023; CESP:C03A7.7
A:Experimental source: strain Bristol NZ, clone C03A7
C:Genetics:
A:Gene: CESP:C03A7.7
A:Map position: 5
A:Introns: 75/3
C:Superfamily: gladiin

```

Oy      2 QEDPQTEC-QCCORRCRQESDPRQ-QQYCRCKETEIE 39
         | : : | | | | | : : | : : | | : :
Db      194 QACQCPCQCQCQCSCVQYQYQYQYQYQYQYQYQYQYQY 233

```

```

RESULT      13
T31889
hypothetical protein C03A7.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T31889
R:Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid C03A7.
A:Reference number: Z21096
A:Accession: T31889
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-438 <GR>
A:Cross-references: EMBL:AF016451; PIDN:AAB65995.1; GSPDB:GN00023; CESP:C03A7.8
A:Experimental source: strain Bristol N2; clone C03A7
C:Genetics:
A:Gene: CESP:C03A7.8
A:Map position: 5
A:Introns: 75/3
A:Superfamily: glladin

```

Query Match	29.8%;	Score 74;	DB 2;	Length 438;
Best Local Similarity	30.2%;	Pred. NO. 1.1;		
Matches 13;	Conservative 12;	Mismatches 12;	Indels 6;	Gaps 2

```

QY      3  EDPQTECQ-QCQRC-----RQESDPQQQYCQRCKEICEE  39
      :  |  ||  |||  :  :  :  :  :  :  :  :  :  :  :
Db     191  QQAQAQCQPQCQCQCQSSCVQQQQPSTQCEPACNTQCSDICQ  233

```

```

RESULT 14
T31898
hypothetical protein C03A7.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T31898
R:Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid C03A7.
A:Reference number: Z21096
A:Accession: T31898
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-445 <GRE>
A:Cross-references: EMBL:AF016451; PIDN:AAB66007.; GSPDB:GN00023; CESP:C03A7.14
A:Experimental source: strain Bristol N2; clone C03A7
C:Genetics:
A:Gene: CESP.C03A7.14
A:Map position: 5
A:Insertions: 18/3; 75/3
A:Superfamily: glnadin

```

Query Match	29.8%;	Score 74;	DB 2;	Length 445;
Best Local Similarity	30.2%;	Pred. NO. 1.1;		
Matches 13;	Conservative 12;	Mismatches 12;	Indels 6;	Gaps 2

```
QY      3 EDPQTETCQ-QCQRRC-----RQESDPRQOQYCQRCKEICEE   39
          : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     191 QQAQAQCQPQCQQQCQSSCVQKQKPSTQCEPACNTQCSDICQ    233
```

```

RESULT 15
T44430
protein PV100 [imported] - winter squash
C:Species: Cucurbita maxima (winter squash)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44430
R:Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin-
A:Reference number: Z22767; MUID:99107919
A:Accession: T44430
A:Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195; NID:q3808061; PTDN:BA434056.1; PID:q3808062

```

	Query Match	29.6%	Score 73.5	DB 2	Length 810
	Best Local Similarity	34.9%	Pred. 1.9		
	Matches 15, Conservative	9	Mismatches 18	Indels 1	Gaps 1
OY	1	NQF-DPQTECCQCCRCRQEQESDPQOQYCCRCRCEICEEEE	42		
db	69	NQFSPRAEYECVRLRCVQAEVGGQKRCCEVCERLRNERQ	111		

Search completed: March 1, 2001, 15:52:37
Job time: 561 sec